Motivation

The Human Rights Watch reported an increase in the harassment of Asians since the start of the COVID-19 pandemic. This rise may have been exacerbated by anti-Asian messages conveyed by prominent world leaders (e.g., former President Donald Trump, former Minister of Education of Brazil Abraham Weintraub). The goals of this research are: (1) to identify patterns of anti-Asian racism on Twitter during COVID-19, (2) build a better understanding of racially-motivated online harassment, and (3) shed light on data patterns applicable to predictive models for identifying instances of racially-motivated online harassment.

Methodology

- **Data Collection** - Twitter API is used and scripted with Python to retrieve tweets that include keywords/hashtags such as: “Chinavirus,” “MakeChinaPay,” “WashTheHate,” “Kungflu,” and “IAmNotAVirus.” Data has been collected from May 26th, 2020 onwards with plans to collect data leading up to the pandemic. The data will be uploaded to MongoDB/Pandas for cleaning, integration, and analysis.

- **Data Analysis** - Aggregation queries will identify properties of keywords such as the number of instances per day, which keywords are most common, and the geographic location (e.g., U.S. state) of anti-Asian tweets. An algorithm developed by Kleinberg will be used to observe temporal patterns of racial harassment. The data will also be explored in relation to the timeline of notable COVID-19-related events.

- **Data Visualization** - Additionally, data visualizations will be generated to help identify and understand connections and insights emerging from the data.

Preliminary Analysis

**Figure 1.** Frequency of racist keywords/hashtags per state. The top 4 states with the highest number of racist keywords/hashtags were Texas (2,953), Florida (2,588), New York (2,480), California (2,385). The 5th highest was Arizona (945).

**Figure 2.** Frequency of negative and positive keywords/hashtags. Keywords/hashtags were categorized as either positive (#WashTheHate) or negative (#Kungflu) and then graphed over the course of several months.

**Figure 3.** Global view of racist keywords/hashtags. The top 2 countries with the highest number of racist keywords/hashtags were India (29,171) and the United States (24,115). The 3rd highest was Canada (3,988).

**Figure 4.** Cumulative flow of COVID-19 keyword/hashtags. The top 3 most common keywords/hashtags were Chinavirus (48.69%), Chinavirus (34.51%), and Kungflu (14.81%). The 4th highest was hateisavirus (0.11%).

**Important Events/Dates**

- **June 20th** – Trump’s rally in Tulsa Calls Covid-19 Coronavirus The ‘Kung Flu’
- **June 23rd** – Trump Once Again Calls Covid 19 Coronavirus The ‘King Flu’
- **July 7th** – The number of coronavirus cases in the U.S. surpassed three million. The United States officially notified the United Nations of its withdrawal from the World Health Organization.
- **July 22nd** – Trump Touts Minority Unemployment ‘Before the China Plague’ to Rebut Biden’s Claim He’s ‘First’ Racist President
- **August 8th** – Trump signs orders aimed at extending some pandemic relief after Congress fails to reach a deal. Trump urges Americans to stop politicizing the coronavirus, blames China.

References


